

```
notausers@one:~$ echo я | awk '/[[:alnum:]]_/'
я
notausers@one:~$ echo я | awk '/^[[:alnum:]]_/'

notausers@one:~$ echo ع | grep -P '[\p{L}]'

notausers@one:~$ echo ع | grep -P '[\p{Arabic}]'

notausers@one:~$ echo ع | grep -P '[\p{Latin}]'

notausers@one:~$ echo عy | grep -P '[\p{Latin}]'

notausers@one:~$ echo لوحة المفاتيح | grep -P '[\p{L}]'

notausers@one:~$ echo لوحة المفاتيح | awk '/[[:alnum:]]_/'

notausers@one:~$ echo ي | awk '/^[[:alnum:]]_/'

http://www.unicode.org/reports/tr18/
http://mlterm.sourceforge.net/

Ну а \w \W пусть остаются для [A-Za-z0-9_]

PS
notausers@one:~$ echo я | awk '/^[^\блабла]_/'
я
:)
```

ع
ع

عy
لوحة المفاتيح
لوحة المفاتيح